

新聞記事との出会いを促す言葉遊びゲームの試作

日置竜輔^{†1} 岩上慎之介^{†1} 田中龍仁^{†1} 保土沢朋和^{†1} 伊藤太一^{†1}
小山内魁人^{†1} 中川翔真^{†1} 金澤快飛^{†1} 服部俊紀^{†1}
寺沢憲吾^{†1} 美馬のゆり^{†1} 角康之^{†1} 坂井田瑠衣^{†1}

概要: 数あるメディアの中でも、新聞には話題の多様性や信頼性、一覧性など多くの利点がある。それにもかかわらず近年は様々なニュースメディアが提供されていることもあり、購読者数は減少している。そこで我々は、北海道新聞社の協力を得て過去 33 年間の新聞記事データを活用し、新聞記事との新しい出会いの場を作り出すべく日常的に遊べる言葉遊びのスマホゲームを試作した。言葉データの素材として新聞記事から年度ごとの特徴語を抽出して利用することで、言葉遊びを通して過去の北海道の時事ニュースやトレンドワードの背景に触れる機会の演出を試みた。被験者実験を行ったところ、ゲームの面白さや没入感は受け入れられその一環で新聞記事への出会いをある程度は達成できた一方で、単語自動抽出に起因した問題や新聞記事への誘導のタイミングについての課題を得た。

1. はじめに

新聞には数あるメディアの中でも話題の多様性や信頼性、一覧性など多くの利点がある。それにもかかわらず近年は様々なニュースメディアが提供されていることもあり、購読者数は減少している。この状況を改善するために我々は、新聞の単語に着目した言葉遊びゲームは新聞記事を読む動機付けに寄与するかどうか研究した。具体的には、北海道新聞社の協力を得て過去 33 年の新聞記事データを活用し、ゲームで使用することができる辞書を作成した。その後、その辞書を有効に使用するために 5 つのゲームを開発した。5 つのゲームでは、それぞれ新聞記事を読むことができ、ゲーム性を活用し新聞との出会いの場を演出している。演出方法の詳細については第 4 章で記載されている。

ゲームを開発した後、ゲームが新聞記事を読む動機付けに寄与するか調査するため実験を行った。実験では実際にゲームをプレイしてもらい、新聞記事に興味をもったか、日ごろ新聞と触れる機会が少ないプレイヤーと新聞記事を引き合わせる有効な手段であったかアンケートを行い調査した。

2. 関連研究

2.1 言語学習へのゲームの活用

これまでに言葉を用いたゲームについての研究は数多く行われてきた。これらの研究で主となっているものは、言語の学習を目的としたものである[1][2]。この目的で研究を行う場合、対象年齢と学習内容を限定していること [1]や対象者が十分に言語を習得していないということ[2]が重要である。

一方で、我々が行った研究では言語の学習を目的にすることは直接的に関わっていない。我々は、新聞記事を読む

動機付けを目的として単語を用いたゲームを開発した。

本研究は、先行研究とは違う目的で研究を進めていくのだが、目的以外でも対象年齢や時代についても先行研究とは異なっている。対象年齢については、新聞に触れる機会が少ない 20 代を対象としている。また我々は、最近の時代の単語や文法のみを使用している先行研究とは異なり、時代性のある単語や文法を使用することでゲームの特徴を増やし、さらに新聞と出会えるような場を演出した。これは、新聞の特徴を伝える大切な要素の 1 つである時代性に着目した結果、特定の年代で制限するのではなく時代性のある単語や文法を使用することが本研究では有効であると想定したためである。

2.2 新聞データからの特徴語の抽出

既存の研究において、新聞データから特徴のある単語を抽出するには様々な手法がとられている。tf-idf を用いた手法[3]や HGS を用いた手法[4]などがある。tf-idf とは、Term Frequency – Inverse Document Frequency の略であり、Term Frequency と Inverse Document Frequency の 2 つの要素を計算することにより単語の重要度を分析する方法である。HGS とは、超幾何分布を応用した確率計算に基づくものであり、その定義から、高頻度語や低頻度語に偏らない、公正な重み付け[5]を可能にする方式である。既存の研究では、文章の要約を行うために各文で特徴のある単語を抽出していた。本研究では形態素解析を中心に単語の出現回数について分析した。

^{†1} 公立はこだて未来大学
Future University Hakodate, Hokkaido, Japan

3. 新聞記事データからのゲーム素材の生成

まず、ゲームを開発する前に新聞記事データをゲームに使用する形にするため分析を行った。今回使用した新聞記事データは北海道新聞社から提供していただいた 1988 年から 2020 年までの CSV データで、記事数は 4,498,758 件である。まず CSV に記載されている記事中に含まれる<>や*といった実際の新聞には掲載されていないが、北海道新聞社内で使用されている文字列を正規表現で除去した。今回作成したゲームには正規表現で整形した文章と名詞が必要なため、名詞の抽出作業を行った。名詞の抽出には形態素解析を行った。形態素解析エンジンには MeCab を使用した。通常の MeCab 内の辞書では「新型コロナウイルス」などといった固有名詞に対応しておらず、抽出できなかったため、mecab-ipadic-NEologd と連携させて辞書の性能向上を図った。その結果、1 つ 1 つの名詞の出現回数を各年代の月ごとに計測することができた。その後、特徴のある名詞を抽出するために、はじめに tf-idf を用いた。tf-idf を用いることで特徴のある名詞を抽出することが可能であった。次に各月の最大値・最小値を比較し差が大きいものを抽出してみた結果、後者の方がより特徴のある名詞を抽出することに成功した。したがって、最大値・最小値を比較し差が大きい単語を、その年の特徴のある名詞（トレンドワード）として抽出し、ゲーム素材として使用した。

4. 試作した 5 つの言葉遊びゲーム

我々は、第 3 章で得られたゲーム素材を用いて 5 つの言葉遊びゲームを試作した。このゲームは、それぞれのゲームのヒントや解説で実際の新聞記事を読むことができ、新聞の特徴を表現しているゲームとなっている。以下に 5 つのゲームの概要を記述する。

Hit And Blow Text

Hit And Blow Text は、Hit And Blow という既存のゲームをもとにひらがなを使って行うゲームである。Hit And Blow とは 4 つの数字のヒントをもとに導いていく数当てゲームのことである。ヒントとして、位置と数字どちらも正しいヒットと、数字は正しいが位置が違うブローの 2 種類が存在する。このゲームでは、数字ではなく新聞記事から抽出した単語のひらがなで行う。設定した単語を当てることができるとその単語の正しい表記、その単語が含まれている新聞記事を読むことができる。このゲームをプレイすることで、本来新聞を読むことで得られる語彙を学習することができ、解説の新聞記事を読むことで単語に関連した文章や単語の使い方を知ることができる。また、1 問あたりにかかる時間が他のゲームと比べ長いので、正解の単語の印象がつきやすいという特徴を持っている。



図 1 Hit And Blow Text のゲーム画面

8 パズル

8 パズルは、3x3 のマスにひらがながランダムに配置され、ひらがなを動かして横一列の 3 文字で単語を作るゲームである。制限時間内に 8 個の単語を作成すること、もしくは制限時間が過ぎることでゲームが終了する。横一列で単語判定をする際に新聞から抽出した単語を使用しているため、プレイヤーが予想した単語以外の 3 文字が単語と判定される場合がある。プレイヤーは今まで知ることがなかった単語に触れることができ、新聞への興味を持たせる工夫がされている。解答後には、プレイヤーが作成した単語に関連した新聞記事を読むことができる。



図 2 8 パズルのゲーム画面

もじもじまぜまぜ

もじもじまぜまぜは、制限時間内にピラミッド状に配置された文字から単語を作るゲームである。一度使用した文字はピラミッドから削除されるという条件下で配置された文字を余すことなく使用し、単語を作るゲームである。ヒントとして、配置された文字から作成できる単語が含まれている記事を見ることができる。ヒントの際に新聞記事を表示することにより、問題となっている単語以外の単語も注意して見てしまうことがこのゲームの特徴となっている。また、ヒントとして表示されるのは問題の単語数により変化し、他のゲームでは結果として閲覧していた新聞記事を、このゲームでは問題を解くと同時に閲覧できるようになっている。



図3 もじもじまぜまぜのゲーム画面

ワードトレンドクイズ

ワードトレンドクイズは、単語と年代の関連性を考える4択クイズである。出題された単語について、4つの年代のうち新聞記事への出現回数が最も多い年代を選択すると正解となる。選択後、各年代の新聞記事への出現回数が表示される。解答後には、出題された単語に関連した新聞記事を読むことができる。このゲームは、新聞の時代性という特徴を有効に使っており、プレイヤーが生まれる前の年代や懐かしいと感じる年代の問題が出題される。この時代性を用いることにより新聞記事との出会いを演出している。



図4 ワードトレンドクイズのゲーム画面

SWAPPING LETTERS

SWAPPING LETTERSは、バラバラに配置された単語の文字を正しい順序に並べ替えるゲームである。ゲーム開始前にプレイヤーは出題される単語の年代を指定する。プレイヤーは自由に文字を移動することができ、正解するか、5回誤答すると次の問題へ進む、または出題単語の記事を読むことができる。10問終了後、出題された単語に関連した新聞記事を読むことができる。このゲームは、指定した年代における流行の単語を活用しているため、プレイヤーは新聞の時代性に触れながらプレイすることができる。また、最終画面で出題された単語の記事を一覧できるという要素がこのゲームの特徴となっている。この特徴があることにより、他のゲームと比べ新聞記事の表示を振り返ることが容易になっている。



図5 SWAPPING LETTERS のゲーム画面

5. 実験

我々は、新聞の単語に着目した言葉遊びゲームは新聞記事を読む動機付けに寄与するかどうかの調査として実験を行った。

今回の実験は、我々が開発したアプリを搭載した実機をプレイしてもらった。この実験は、2021年12月14日から2021年12月20日にかけて計4回行った。被験者はアルバイト募集に応募してきた公立はこだて未来大学の学生である。

被験者からのアンケート調査の結果、「ゲームは面白かったですか?」という質問に対して、10人中9人の被験者が「面白い」と回答した。また、「このゲームを通して新聞記事に興味を持ちましたか?」という質問に対して、10人中8人の被験者が「興味を持った」と回答した。その他に被験者からの意見として多かったのが、「ゲームの動作が重い」、「難易度が高いので低くしてほしい」などの意見が出た。

今回の実験は被験者が10名であったため、詳細な評価は難しい状況であった。しかし、ゲームの面白さや没入感は受け入れてもらった。アンケート結果からもわかるように、言葉遊びゲームは面白い要素があることや、新聞記事と日ごろ新聞と触れる機会が少ないプレイヤーを引き合わせることに一定程度有効的であることが確認できる。一方、言葉遊びゲームを通して時代の特徴を知ることにはあまりできていないという結果も得られたため、各ゲームの時代と新聞との関連性が低いと予想される。そのため、改善が必要であると考えられる。

6. おわりに

本研究では、新聞記事データを分析、加工することでゲームに使うための素材を作った。今回のデータ整形に当たって、データの抽出を行う際に tf-idf を用いるより各年の頻度の最大値と最小値を比べた方が、精度の向上に繋がったことを紹介した。しかし、この手法を用いると新聞記事を

執筆した記者の名前や、写真を撮影した人物名を抽出する問題が生じた。誤って抽出してしまった記者の名前に関しては、隅付き括弧【】で囲まれていることが多かったため、これを取り除くことが今後の課題となる。

また、言葉遊びに着目した5つのゲームの改善点としてロード時間を短縮することと、問題として提示した単語を新聞記事で表示するときにハイライトで表示するということがプレイヤーのレビューからあがった。ロード時間については、加工した新聞記事すべてのCSVファイルを読み込んでいるため長時間かかってしまう。この課題を解決するには、必要最低限のCSVファイルのみ読み込むこととCSVファイルの処理のアルゴリズムを改善することが挙げられる。今回文章をデータ化していることもあり、改行ありのセルを適切に抽出することに苦労した。その時に用いた手法は全探索をし、ダブルクォーテーションを探し出し分割していくというものであった。これを改善するために、尺取り法などの計算量の少ないアルゴリズムに変更する必要がある。CSVファイル以外でも全探索の方法をとってきたため、アルゴリズムの改善はロード時間の短縮につながると想定した。

そして、このゲームは新聞記事を読めることができ、ユーザと新聞を引き合わせる有効な手段であると考え、実験を行った。結果として、被験者からのアンケート調査の結果、「ゲームは面白かったですか?」という質問に対して、10人中9人の被験者が「面白い」と回答した。反対に「まったく面白くない」と回答した被験者は理由として、「無理矢理、新聞とゲームに関連付けた気がしたから」と記入していた。このような意見が出ているため、ゲームの見た目や機能を新聞と関連付ける必要があることが分かった。今回実験に参加した被験者は10人と少なく、新聞記事を読む動機付けに寄与するかどうか評価することは厳しいものであった。しかし、評価内容としては肯定的なものが多かったため、レビューに挙げられている課題を改善させていき再度実験を行うことで新聞記事を読む動機付けに寄与するかどうか評価することは可能であると結論付けた。

謝辞

本プロジェクトでは北海道新聞の過去33年分の新聞記事データを活用した。プロジェクトの構想やシステム試作において北海道新聞函館支社長の三浦辰治氏に多大なるご協力を頂戴したので感謝申し上げます。

参考文献

- [1] 加藤映子, 語学教育とゲーム, コンピュータ&エデュケーション, 2011, vol.31, no. 1, pp. 28-33.
- [2] 阿部征一郎, 言葉遊びと語彙学習, Artes Liberales, 1991, no. 49, pp. 54-62.
- [3] T. Nakashima, R. Nakamura, Daily clustering for the electronic newspaper based on the analysis of trends, IEEE Pacific Rim Conference on Communications, Computers and Signal

Processing, 1999, no. 99.

- [4] 多田知道, 岩沼宏治, 鍋島英知, F-047 イベント時系列マイニングを目的とする新聞記事からの時系列情報に基づく単語抽出情報科学技術フォーラム講演論文集, vol. 7, 2008, no. 2, pp. 425-428
- [5] 久光徹, 丹波芳樹, 組み合わせ的確率モデルに基づく特徴単語選択方法 -超幾何分布の応用-, 情報処理学会研究報告自然言語処理, 2000, vol. 2000, no. 107, pp. 85-90.